Volume 1, Issue 5 December 2015 ISSN NO: 2454 - 7514

North Asian International Research Journal Consortium

North Asian International Research Journal Of Science, Engineering and Information Technology

NAIRJC JOURNAL PUBLICATION

North Asian International Research Journal Consortium

Welcome to NAIRJC

ISSN NO: 2454 -7514

North Asian International Research Journal of Science, Engineering & Information Technology is a research journal, published monthly in English, Hindi, Urdu all research papers submitted to the journal will be double-blind peer reviewed referred by members of the editorial board. Readers will include investigator in Universities, Research Institutes Government and Industry with research interest in the general subjects

Editorial Board

Address: - Dr. Ashak Hussain Malik House No. 221 Gangoo, Pulwama, Jammu and Kashmir, India - 192301, Cell: 09086405302, 09906662570, Ph. No: 01933-212815,
Email: nairjc5@gmail.com, nairjc@nairjc.com, info@nairjc.com Website: www.nairjc.com

SPEECH RECOGNITION USING AUTOMATED MFCC BASED DEREVERBATION OF NOISY SPEECH SIGNALS

JASMINE*
*M.Tech Student (ECE Department), Chandigarh Engineering College, Landran, Mohali

SAKSHI SHARMA**
**Assistant Professor (ECE Department) Chandigarh Engineering College Landran, Mohali

ABSTRACT

The proposed work is based on speech derevebation and ASR system. The paper contains a brief review about speech reverberation effect removal techniques with their merits and demerits. The second section contains various works done by authors on speech derevebation. The third section contains the current issue dealing with ASR (Automatic Speech Recognition) system. The fourth section contains the evaluation of the proposed MFCC based dereverbent speech recognition; the proposed system has been tested on data base created using live recording of different voice samples. In the final section these results of the system performance have been evaluated on the basis of PSNR and WER.

Keywords- MFCC, ASR, DEREVERBATION, PSNR, WER

INTRODUCTION

Reverberation is a wonder in amphitheaters, for example, show corridors and chapels. Reverberation comprises of a mix of different echoes, and its force and length of time relies on upon elements, for example, the measurements of the fenced in area, materials utilized as a part of development and shape. Reverberation is alluring in music multiplication, then again, it renders discourse confused. In this way there is a necessity to control resonation of discourse. Clean discourse securing from inaccessible amplifiers can include single or various receivers. A discourse sign caught utilizing a solitary far off receiver is spread because of resonation. Resonation is a wonder in which weakened and deferred forms of a sign are added to itself. The resonation corrupts the nature of discourse, makes sound indiscernible and decreases the exactness of discourse acknowledgment based frameworks. When all is said in done, the reverberant discourse sign can be displayed as convolution of the perfect discourse signal and the acoustic room motivation reaction.

A dereverberation calculation forms the watched discourse signal, in order to frame an evaluation of the spotless discourse signal by figuring the acoustic room motivation reaction much of the time. Henceforth, the technique for discourse dereverberation is seen as a visually impaired deconvolution issue as neither the perfect discourse signal nor the AIR is accessible by and large. Different calculations have been proposed for discourse dereverberation in this connection [1], [2].

Blind deconvolution techniques perform viable derever-beration however are hard to execute for all intents and purposes as they have high computational multifaceted nature and high affectability to commotion. In [3], the dereverberation is done by utilizing cepstrum to focus the AIR and after that utilization backwards sifting to acquire the evaluation of clean discourse. The truncation mistake present in [3] was uprooted in [4], yet at the same time reverse separating was needed. In this paper, a technique which makes utilization of complex cepstrum and direct forecast (LP) lingering sign to deconvolve the resonated discourse sign is proposed. This technique gauges clean discourse without really evaluating the AIR and in this way evades the computational multifaceted nature included in converse sifting.

The impact of reverberations in a room on discourse signs is a basic issue in numerous discourse applications. For instance, it is vital to wipe out reverberations if we are to accomplish hearty programmed discourse acknowledgment (ASR) in genuine situations. Reverberation in rooms extremely changes signal attributes, and in this manner debases acknowledgment execution. Much exertion has been committed to the deriverberation issue utilizing both single and various channel based strategies [5–15], yet no attractive strategy has been discovered yet. With respect to single channel dereverberation, a method has been created for evaluating the reverse channel of a room exchange capacity utilizing the consonant structure of discourse [5,6]. It functions admirably for long reverberation times, however functional utilization is still constrained because of the substantial measure of discourse information needed. Another single amplifier technique [7] proposes improving discourse districts where direct discourse parts are overwhelming contrasted and the reverberant parts of the signal.

The sign quality in different discourse correspondence applications, for example, video chatting, hands free telephony, and voice-controlled frameworks is traded off from multiple points of view. A first sort of aggravation is the supposed acoustic echoes, which emerge at whatever point an amplifier sign is gotten by the microphone(s). A second wellspring of sign disintegration is clamor and unsettling influences that are added to the sign of interest.Finally, extra flag debasement happens when reverberation is added to the sign as it spreads through the recording room reflecting off dividers, questions, and individuals. This spread results in a sign constriction and phantom twisting that can be displayed well by a direct channel. Nonlinear impacts are normally of second-request and fundamentally originate from the nonlinear attributes of the amplifiers. The straight channel that relates the transmitted sign to the got sign is known as the acoustic motivation reaction [16] and assumes an imperative part in numerous sign upgrade methods. Frequently, the acoustic drive reaction is a non minimum stage framework, and can in this manner not be causally altered as this would prompt an insecure acknowledgment. In any case, a non causal stable reverse may exist.

Whether the drive reaction is a minimum stage framework relies on upon the resonation level. the state of the recording room and the position of source and audience. Next come an arrangement recently reflections, likewise called resonation, which rot exponentially in time. These driving forces stem from multipath proliferation as acoustic waves reflect off dividers and items in the recording room. As articles in the recording room can move, acoustic motivation reactions are commonly exceedingly time-shifting. Despite the fact that flags (music, e.g.) may sound more lovely at the point when resonation is included, (particularly for discourse flags), the comprehensibility is ordinarily lessened. With a specific end goal to adapt with this sort of misshapening, dereverberation or deconvolution procedures are called for.Whereas upgrade methods for acoustic reverberate and clamor lessening are understood in the writing, high caliber, computationally effective dereverberation calculations are, to the best of our insight, not yet available.

II. LITERATURE REVIEW

[17] Stephen T. Neely et al., Inverse Filtering - Least squares method gives a Spin-off of framework distinguishing proof, there is a prerequisite to acquire the backwards of the recognized channels. It can be normal that the opposite channel can be gotten by the scientific reversal of the motivation reaction, be that as it may, on account of room drive reactions, the opposite channel will be unsteady. This is on the grounds that room motivation reactions are for the most part non least phase.

[18] M. H. (Monson H.) Hayeset al., The NLMS calculation is one of a class of calculations, which gives an estimate of the Wiener answer for a given framework setup. In the first piece of this work, we utilized the NLMS calculation as portrayed in to acquire a regulated opposite channel of an acoustic channel.

[19] Yiteng Huanget al., Frequency Domain Normalized Multichannel LMS. The inspiration for the recurrence area standardized multichannel LMS calculation (FNMCLMS) is displayed in two stages. Initially the recurrence space unnormalized multichannel LMS calculation is determined. Besides, the calculation is altered for standardization. Standardization of the calculation is fancied in light of the fact that the normalized recurrence space multichannel LMS has a moderate joining rate. The moderate joining is expected to the cross coupling between the channels, the general meeting rate is controlled by the slowest merging channel

Newton's strategy is connected in the standardization system; further, rough guesses are used to rearrange the overhauled calculation. Accordingly eigenvalue contrasts are lessened and joining is progress

[20]Jont B. Allenet al., Algorithm performance with higher order impulse responses whereTime area calculations were computationally serious with longer channel drive reactions. The time area multichannel newton and recurrence space standardized multichannel newton calculations were seen to gauge the channel drive reaction. The center was on the recurrence area standardize multichannel LMS. Reenacted room motivation reactions were likewise used to watch the execution of the calculations. The room motivation reaction generator, utilizing the picture method, with the usage in was utilized to acquire the room motivation reactions.

PROBLEM STATEMENT

In the current research the issue is to reduce the reverb effect in the sound recorded with different mics this induces the reverb effect at different level, so a standardized effect removal is needed which can adjust to the changing frequency rate of the recorded sound samples and downscale the attenuation provided when a signal is filtered it also reduces a few details which can be an issue when the de-reverberation is followed by ASR system.

PROPOSED SYSTEM

Collect dataset for the recognition system. Select the speech file for processing from database. Add noise to speech files to degrade the speech features. Denoise the speech files using supervised learning of noise and reverb pattern based on MFCC feature change using reverb harmonic calculation. Again analyze the features of the denoised sound files and create feature map for all the files using cepstral technique. Again insert a query sound file for matching with the created database of feature files using HMM feature distance by Euclidian distance of hidden features extracted in the 3 step, the WER is calculated for the retrieved sound file and feature match accuracy is determined by

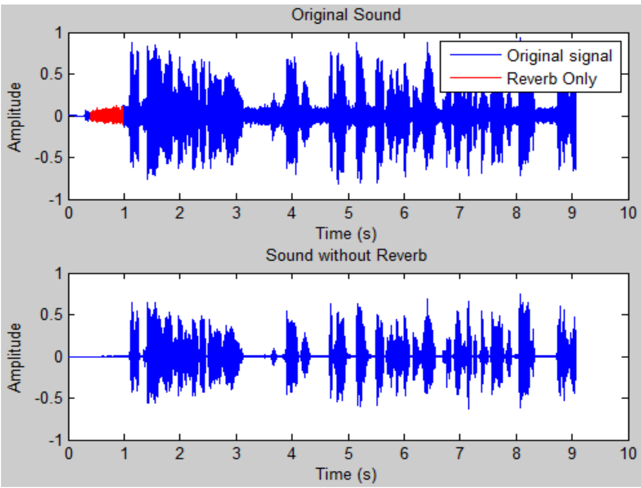

Figure 1 shows the original sound file without reverberation and the sound file as reverb file

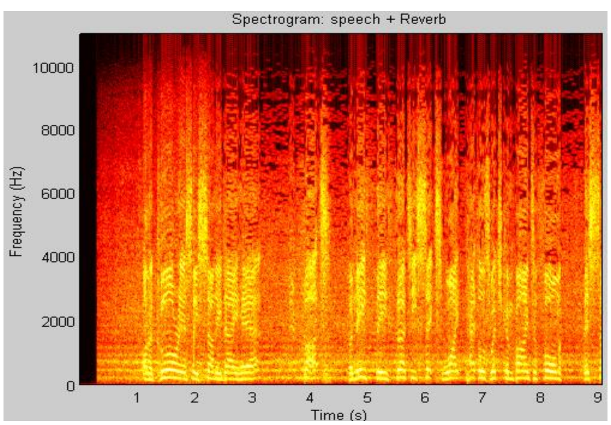

Figure 2 shows the reverb pattern in a reverbant sound file in spectrogram

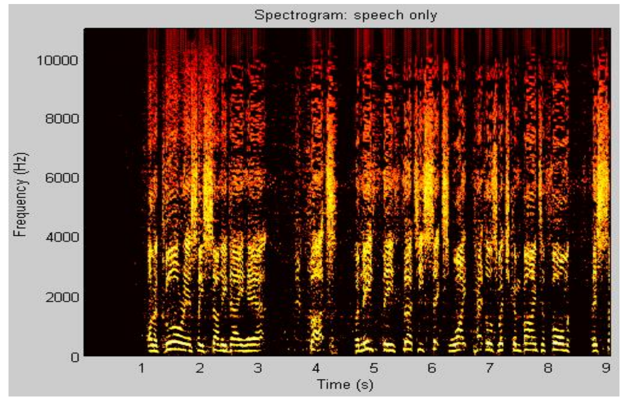

Figure 3 shows the output of the proposed system of de-reverberation

The following section is the speech recognition section which shows the accuracy of the ASR feature based HMM-MFCC classifier foe match accuracy precision and recognition of the de-reverb sound files

RESULTS AND DISCUSSION

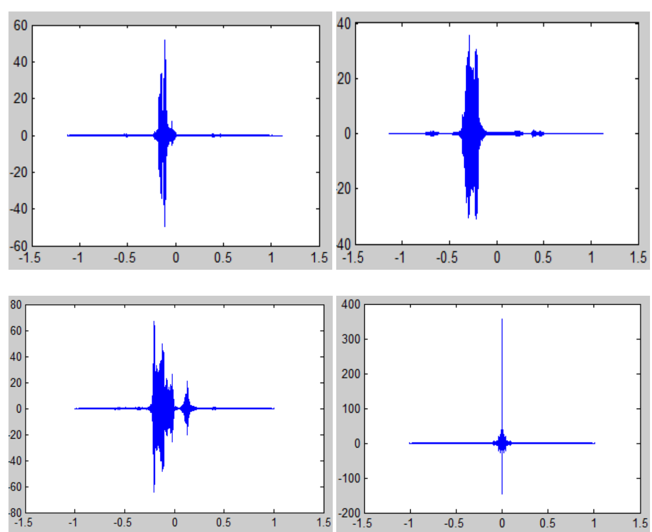

Figure 4 shows the dereveb sound files of the Punjabi letter *UDDA* from four users showing high variation.

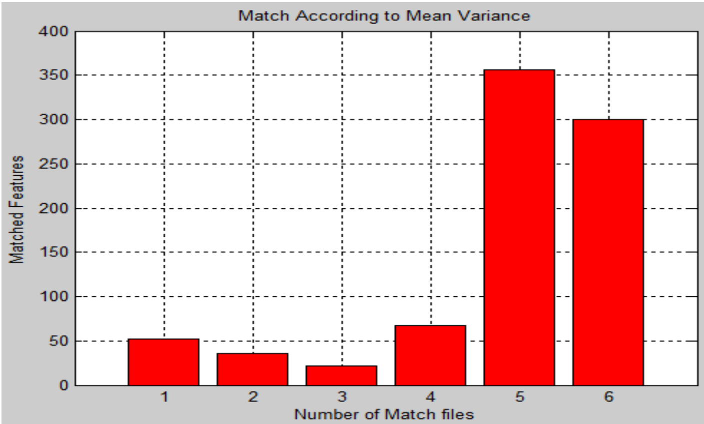

Figure 5 shows the output of the ASR system returning query output sounds, showing high match in feature outputs the system accuracy has increased with proposed dereveb system as a result following table shows the WER for user 1 with four different mics:

Table 1 shows the performance output of the proposed system

	Mic system	Average PSNR	WER percentage
User 1	**Mic 1**	**30 db**	**30 %**
	Mic 2	**27 db**	**35%**
	Mic 3	**25.6 db**	**40%**
	Mic 4	**31 db**	**25%**

CONCLUSION

The proposed research work is based on the concept of removing noise from live recorded speech in isolation format using different impedance mics, by different users, the study of reverb removal is done to enhance the degraded quality of the sound file, this dereverbation uses the MFCC component analysis, this system is robust to the noisy reverb which is mixed form of reverbation and reduced original signal with high density attack after which the ASR shows low match accuracy due to change in the speech/sound file pattern, in future more samples will be combined and different users will be analyzed for pitch change and frequency change which will be measured with the time, WER, improvement in SNR values and correlation between the original and dereveb sound.

REFERENCES

[1]S. Subramaniam, A.P. Petropulu, and C. Wendt, "Cepstrum-based de-convolution for speech dereverberation," *Speech and Audio Processing,IEEE Transactions on*, vol. 4, no. 5, pp. 392–396, 1996.

[2]RA Kennedy and BD Radlovic, "Iterative cepstrum-based approach for speech dereverberation," in *Signal Processing and Its Applications,1999. ISSPA'99. Proceedings of the Fifth International Symposium on*.IEEE, 1999, vol. 1, pp. 55–58.

[3]D. Bees, M. Blostein, and P. Kabal, "Reverberant speech enhancement using cepstral processing," in *Acoustics, Speech, and Signal Processing,1991. ICASSP-91., 1991 International Conference on*. IEEE, 1991, pp.977–980.

[4]S. Xizhong and M. Guang, "Complex cepstrum based single channel speech dereverberation," in *Computer Science & Education, 2009.ICCSE'09. 4th International Conference on*. IEEE, 2009, pp. 7–11.

[5] T. Nakatani, M. Miyoshi and K. Kinoshita, ''One microphone blind dereverberation based on quasi-periodicity of speech signals,'' in Advances in Neural Information Processing Systems 16 (NIPS16) (to appear), (MIT Press, Cambridge, Mass., 2004).

[6] T. Nakatani and M. Miyoshi, ''Blind dereverberation of single channel speech signal based on harmonic structure,'' Proc.ICASSP '03, Vol. 1, pp. 92–95 (2003).

[7] B. Yegnanarayana and P. S. Murthy, ''Enhancement of reverberant speech using LP residual signal,'' IEEE Trans. Speech Audio Process., 8, 267–281 (2000).

[8] M. Unoki, M. Furukawa, K. Sakata and M. Akagi, ‘‘A method based on the MTF concept for dereverberating the power envelope from the reverberant signal,’’ Proc. ICASSP ’03, Vol. 1, pp. 840–843 (2003).

[9] R. Roy and T. Kailath, ‘‘ESPRIT: Estimation of signal parameters via rotational invariance techniques,’’ IEEE Trans. Acoust. Speech Signal Process., 37, 984–995 (1989).

[10] O. R. Schmidt, ‘‘Multiple emitter location and signal parameter estimation,’’ IEEE Trans. Antennas Propag., 34, 276–280 (1986).

[11] H. L. Van Trees, Optimum Array Processing, PartIV ofDetection, Estimation, and Modulation Theory (John Willey & Sons, New York, 2002).

[12] M. Miyoshi and Y. Kaneda, ‘‘Inverse filtering of room acoustics,’’ IEEE Trans. Acoust. Speech Signal Process., 36, 145–152 (1988).

[13] S. Amari, S. C. Douglas, A. Cichocki and H. H. Yang, ‘‘Multichannel blind deconvolution and equalization using the natural gradient,’’ Proc. IEEE Workshop on Signal Processing in Advances in Wireless Communications, Paris, pp. 101–104,April (1997).

[14] X. Sun and S. C. Douglas, ‘‘A natural gradient convolutive blind source separation algorithm for speech mixtures,’’ Proc. ICA ’01, pp. 59–64 (2001).

[15] R. Aichner, S. Araki and S. Makino, ‘‘Time domain blind source separation of non-stationary convolved signals by utilizing geometric beamforming,’’ Proc. NNSP ’02, pp. 445– 454 (2002)

[16] H. Kuttruff, *Room Acoustics*, Applied Science Publishers, Essex, England, 2nd edition, 1979.

[17] Stephen T. Neely and Jont B. Allen. Invertibility of a room impulse response. Journal of the Acoustic Society of America 66(1), VOL. 66(1):165–169, JULY 1979.

[18] M. H. (Monson H.) Hayes. Statistical digital signal processing and modeling. 1996.

[19] Yiteng Huang, Jacob Benesty, and Jingdong Chen. Acoustic MIMO Signal Processing (Signals and Communication Technology). Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[20] Jont B. Allen and David A Berkley. Image method for efficiently simulating smallroom acoustics. Journal of the Acoustical Society of America, 65(4):943–950, 1979

Publish Research Article

Dear Sir/Mam,

We invite unpublished Research Paper,Summary of Research Project,Theses,Books and Book Review for publication.

Address:- Dr. Ashak Hussain Malik House No-221, Gangoo Pulwama - 192301
Jammu & Kashmir, India
Cell: 09086405302, 09906662570,
Ph No: 01933212815
Email:- nairjc5@gmail.com, nairjc@nairjc.com , info@nairjc.com
Website: www.nairjc.com